# Effect of Spatial Prediction Method on Determination of Anomalous Disease Clusters

Jian Xing[1], Howard Burkom[2], Michael Leuze[3], James Edgerton[2], John Copeland[1], Steve Bloom[3], and Jerome Tokars[1]

[1]Centers for Disease Control and Prevention, [2]The Johns Hopkins Applied Physics Lab, and [3]Science Applications Incorporated

## Objective

Apply spatial scan statistics to data from CDC's BioSense system to:

- Examine the effect of the spatial prediction method on determination of anomalous disease clusters.
- Decide on a reliable spatial estimation method for one BioSense data source.
- Establish criteria for making this decision using other sources.

## Background

Determination of anomalous spatial and temporal disease clusters is a crucial part of BioSense daily disease monitoring task. While many studies have focused on improved computation time and more general cluster shapes, our effort focused on finding anomalies that are correct according to available BioSense data history. Using a newly developed computer program implementing scan statistics, we compared three spatial estimation methods on BioSense data.

## Methods

- Outpatient diagnoses from 32 Department of Defense clinics in or near Texas.
- Study period 1/1/2004 –12/31/2006.
- Daily visit count data for respiratory and rash syndromes studied, for both a rich, clinically broad time series and a sparse series.
- Cluster analysis done at the facility level.
- Applied goodness-of-fit statistic to show how well each prospective estimation method predicts the observed case distribution.
- Goodness-of-fit statistic = the mean of $\frac{(\text{observed} - \text{expected})^2}{\text{expected}}$
- Developed a computer program written in the C language using the circular cluster search technique, Poisson log likelihood detection statistic, and Monte Carlo significance testing. This program also allowed implementation and adjustment of various spatial estimation methods.
- Utilized the program to find all one-day clusters with p-value $\leq$ 0.003 and including at least 2 facilities, using three estimation methods.
- Method one – simple baseline mean: sliding baseline averages with a 2-day buffer between baseline and test day. For the one-week baseline, this method is similar to the EARS C2 algorithm[1]. Named C2_7, C2_28 and C2_56 for one, four and eight -week baselines.
- Method two – baseline mean stratified by weekday/non-weekday with a 2-day buffer between baseline and test day. Named W2_28 and W2_56 for four and eight-week baselines.
- Method three – conditioning on marginal totals for both day and facility as in the SaTScan space-time permutation method[0]. This method includes the test day counts in the baseline and does not use a buffer interval. Named STM_7, STM _28 and STM _56 for one, four and eight-week baselines.

## Results

- Clusters considered significant were those covering at least 2 facilities with p-value $\leq$ 0.003. An example of a one-day cluster found on January 3, 2006 is shown in Figure 2.
- Rash syndrome: mean count 2.69 records per facility per day (3.57 per weekday, 0.51 per weekend).
- Respiratory syndrome: mean count 24.66 records per facility per day (31.76 per weekday, 6.96 per weekend).
- The Chi-Square goodness-of-fit statistic test (Table 1):
  - (i) The STM method provides a better fit than C2 and W2 method on weekday, as expected from the built-in bias of including the current cases in the baseline.
  - (ii) W2 method provides much better fit than C2 and STM on non-weekday, as expected visit counts on non-weekday different from weekday.
  - (iii) In general, the shorter baseline provides a better fit than longer baselines on weekday.
  - (iv) These findings suggest that for a syndrome with a seasonal effect, such as Respiratory, a longer baseline can worsen the fit of the expected spatial distribution, In such situations, it may be possible to model the time series of counts from individual subregions, depending on the data set.
- The clusters were calculated by each category (Table 1):
  - (i) The cluster counts were higher as baseline increased for Respiratory, while one-week baseline had highest cluster count for Rash, additional evidence for the effect of seasonality.
  - (ii) The mean facilities per cluster were higher as baseline increased for Respiratory and decreased for Rash.
  - (iii) The finding shows additional evidence for the effect of seasonality.

Table 1: Goodness-of-fit and Cluster Found by Spatial Estimation Method and Baseline Duration *

| | Method-Baseline Duration | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Method | C2 | | | W2 | | STM | | |
| Baseline | 7 | 28 | 56 | 28 | 56 | 7 | 28 | 56 |
| **Rash** | | | | | | | | |
| Goodness-of-fit on weekday | 1.33 | 1.25 | 1.3 | 1.19 | 1.25 | 0.89 | 1.09 | 1.17 |
| Goodness-of-fit on weekend | 0.8 | 0.75 | 0.72 | 0.52 | 0.49 | 0.75 | 0.72 | 0.71 |
| Number of Clusters | 15 | 6 | 12 | 7 | 5 | 13 | 8 | 10 |
| Mean Clusters per facility | 0.5 | 0.2 | 0.4 | 0.2 | 0.2 | 0.4 | 0.3 | 0.3 |
| Mean Clusters per 100 days | 1.4 | 0.5 | 1.1 | 0.6 | 0.5 | 1.2 | 0.7 | 0.9 |
| Mean Clusters per 100 weekday | 1.5 | 0.5 | 1.3 | 0.6 | 0.5 | 1.4 | 0.8 | 1.0 |
| Mean Clusters per 100 weekend | 1.0 | 0.6 | 0.6 | 0.6 | 0.3 | 0.6 | 0.6 | 0.6 |
| Mean facilities per cluster | 3.4 | 3.0 | 2.7 | 3.4 | 3.2 | 2.7 | 2.9 | 2.7 |
| **Respiratory** | | | | | | | | |
| Goodness-of-fit on weekday | 2.58 | 2.52 | 2.82 | 2.51 | 2.96 | 1.71 | 2.09 | 2.38 |
| Goodness-of-fit on weekend | 3.69 | 3.53 | 3.51 | 1.6 | 1.86 | 3.6 | 3.44 | 3.43 |
| Number of Clusters | 205 | 208 | 225 | 109 | 145 | 175 | 187 | 220 |
| Mean Clusters per facility | 6.4 | 6.5 | 7.0 | 3.4 | 4.5 | 5.5 | 5.8 | 6.9 |
| Mean Clusters per 100 days | 18.7 | 19.0 | 20.5 | 9.9 | 13.2 | 16.0 | 17.1 | 20.1 |
| Mean Clusters per 100 weekday | 18.4 | 18.7 | 20.7 | 11.5 | 16.0 | 14.5 | 15.7 | 20.2 |
| Mean Clusters per 100 weekend | 19.4 | 19.7 | 20.1 | 6.1 | 6.4 | 19.7 | 20.4 | 19.7 |
| Mean facilities per cluster | 3.4 | 3.6 | 3.7 | 4.1 | 4.2 | 3.5 | 3.6 | 3.8 |

* Clusters considered significant were those covering at least 2 facilities with p-value $\leq$0.003 calculated by flat baseline average (C2), stratified baseline average (W2), and conditioning on space & time marginals (STM), for multiple baseline lengths.

Figure 1: The Map of 32 DoD Facilities Inside and Near Texas.

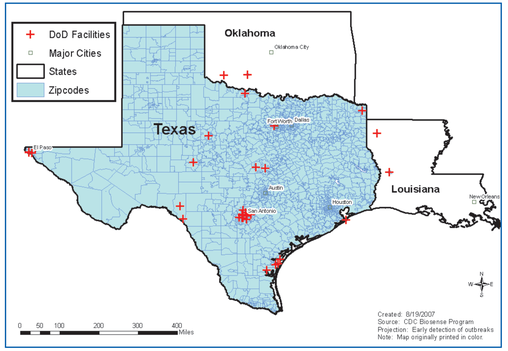
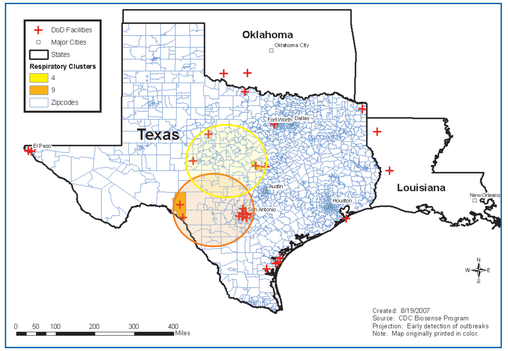



Figure 2: Two Respiratory Clusters Found on January 3, 2006
Cluster One Included 4 Facilities, Cluster Two Included 9 Facilities.




## Conclusions

- Since there is no gold standard, we do not know how many clusters should be found. However, since identification of clusters of public health significance by these methods is rare, methods that find fewer clusters may be preferred.
- Population health monitors need to find unusual disease clusters based on available BioSense data. The clusters should be based on true unusual case location distributions, not irrelevant changes in the data. If there are too many clusters, the system will not be used.
- For DoD BioSense datasets as used for this study, the usual distribution differs from the general population census distribution for several reason:
  - (i) For BioSense data sources in general, population distributed according to those eligible for medical care and care providers contributing to the data source.
  - (ii) Military and civilian population spatial distributions differ, especially given the high concentrations of military personnel, dependents, and retirees in certain regions.
  - (iii) No up-to-date source of military population data available for online adjustment.

Therefore, the method for estimating the distribution is very important for finding significant clusters.

- Accounting for the day-of-week effect and similar systematic features can sharply reduce the number of statistically significant of unknown epidemiological importance clusters (Table 1).
- A longer baseline may not improve the spatial distribution estimation. In this study, an 8-week baseline appeared to be too long for the Respiratory syndrome, probably because of seasonal spatial changes in the data.
- The weekday/non-weekday stratified baseline averaging method (W2) gave the lowest cluster rates in this study, and most effective baselines were 28 and 56 days for the Respiratory and Rash syndromes, respectively.
- The goodness-of-fit statistic is correlated to the number of significant clusters identified and can be used to assess the utility of scan statistics and to choose thresholds, baseline lengths, and other parameters affecting surveillance design.
- Optimal methods for calculating expected values may differ by data source and syndrome, i.e., Rash and Respiratory have much different case counts, mean, trends, seasonality, etc.
- Further study will include additional geographic areas, data types, new statistic algorithms, as well as statistical signal injection and detection for different data pattern. Critical examination of the results will help to establish by context the best methods for calculating expected values for spatial cluster determination.

REFERENCES

[1] Hutwagner L, et al. The Bioterrorism Preparedness and Response Early Aberration Reporting System (EARS). J Urban Health. 2003 ;80 1:89-96.

[2] Kulldorff M, et al. (2005) A space-time permutation scan statistic for the early detection of disease outbreaks. PLoS Medicine, 2:216-224.

ACKNOWLEDGEMENT

Special thanks to James Tobias for great help with GIS technology to contruct maps for cluster visualization.